

UNCLASSIFIED

AD 298 197

*Reproduced
by the*

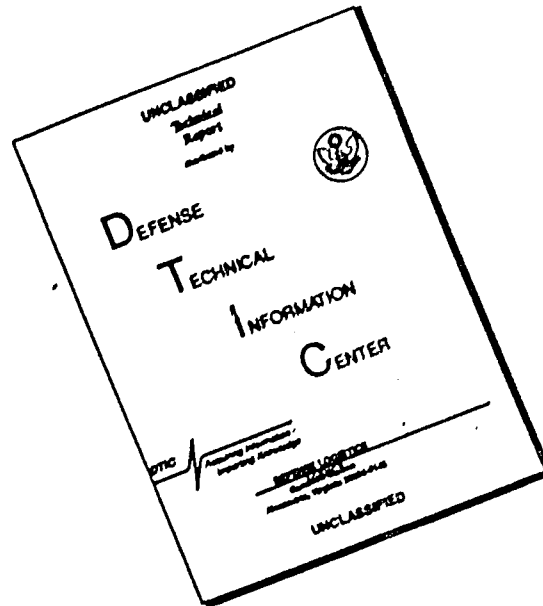
**ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA**



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

63-2-5

CATALOGUE
AS AD NO 298197
ASTIA

A GEOMETRIC INTERPRETATION OF LAGRANGE MULTIPLIERS

R. C. Kao

February 1963

ASTIA
MAR 18 1963
TISIA

A GEOMETRIC INTERPRETATION OF LAGRANGE MULTIPLIERS*

R. C. Kao**

The RAND Corporation, Santa Monica, California

A fundamental assumption common to all economic analyses is the maximization or minimization of an objective function (representing, say, utility, cost, welfare or the like) subject to certain constraints. Statements of the type: "A consumer with given income maximizes his total utility only if his marginal utilities for the various commodities are proportional to their prices," are almost commonplace in economic texts and are generally described as "equilibrium conditions" of the economic process under consideration. Nevertheless, when these meaningful economic theorems are presented to even the more advanced students, the argument is usually shrouded with a complete or partial mystery around the so-called Lagrange multipliers. Very little explanation is given to these multipliers themselves except that they are the coefficients used to form a certain Lagrangian function, the extremization of which leads to the

*I am indebted to Professor A. A. Alchian for calling my attention to this problem.

**Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff.

desired equilibrium conditions.* This paper is devoted to a pedagogic clarification of the intrinsic meaning of these multipliers themselves and a natural reformulation of the equilibrium conditions which permits a better insight into the nature of constrained extremum problems in economics.

Let

$$(1) \quad y = f(x)$$

be a real-valued function of a single variable x . The function f may represent the short-run cost curve of a production process with only one variable factor. If f is sufficiently smooth (i.e. x is infinitely divisible) a necessary condition for a (relative) minimum of (1) is, as is well known,

$$(2) \quad \frac{dy}{dx} = f'(x) = 0,$$

and a sufficient condition for a (relative) minimum of (1) is (2) plus

$$(3) \quad \frac{d^2y}{dx^2} = f''(x) = \frac{d}{dx} f'(x) > 0.$$

* Cf. inter alia the following well known economic texts: R. G. D. Allen, Mathematical Analysis for Economists, London: MacMillan, 1949, pp. 366-367; idem, Mathematical Economics, London: MacMillan, 1956, pp. 610, 614; D. W. Bushaw and R. W. Clower, Introduction to Mathematical Economics, Homewood, Illinois: Irwin, 1957, p. 331; J. M. Henderson and R. E. Quandt, Microeconomic Theory, A Mathematical Approach, New York: McGraw-Hill, 1958, pp. 273-274; J. R. Hicks, Value and Capital, 2nd ed., London: Oxford, 1946, p. 305; P. A. Samuelson, Foundations of Economic Analysis, Cambridge, Massachusetts: Harvard, pp. 362-365; and Taro Yamane, Mathematics for Economists, Englewood, New Jersey: Prentice-Hall, 1962, pp. 116-123.

Geometrically, (2) states that the tangent vector to the curve C defined by (1) must be horizontal; and (3) states that it is increasing in slope around any root x^0 of (2). (Figure 1)

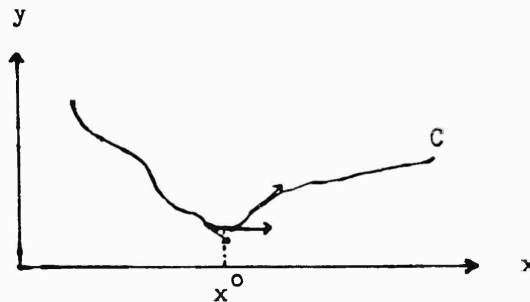


Figure 1

A more easily generalizable geometric interpretation of (3) is the following: Any function f possessing sufficient number of derivatives (i.e., sufficiently smooth) may be expanded into a Taylor's series:

$$f(x) = f(x^0) + \frac{df}{dx}(x - x^0) + \frac{1}{2} \frac{d^2f}{dx^2}(x - x^0)^2$$

$$(4) \quad + \dots + \frac{1}{n!} \frac{d^nf}{dx^n}(x - x^0)^n + \dots$$

where all derivatives are to be evaluated at x^0 . That is, the value of f at x may be represented by its value at x^0 , together with all derivatives of f at x^0 . Consequently, if x^0 is to be a relative minimum, all sufficiently close neighboring x must not yield a smaller $y = f(x)$, i.e.

$$(5) \quad f(x) - f(x^0) \geq 0,$$

or in terms of (4),

$$(6) \quad \frac{1}{2} \frac{d^2 f}{dx^2} (x - x^0)^2 \geq 0$$

since at x^0 , $\frac{df}{dx} = 0$; and if x is sufficiently close to x^0 , the term shown in (6) will dominate the combined effect of all other terms in the expansion (4) because all remaining terms involve $x - x^0$ to a higher order. That (6) is equivalent to (3) is obvious.

If f is now a function of two independent variables, (1) may be rewritten as

$$(7) \quad y = f(x_1, x_2)$$

and a pair of necessary conditions corresponding to (2) are

$$(8) \quad \frac{\partial y}{\partial x_1} = f_{x_1} = 0, \quad \frac{\partial y}{\partial x_2} = f_{x_2} = 0$$

These conditions state that the tangent vectors to the surface S_{12} defined by (7) in the directions of increasing x_1 and x_2 must be horizontal, that is, parallel to the $x_1 x_2$ plane. (Figure 2) If f is sufficiently smooth, its Taylor expansion around any root x^0 of (8) is given by

$$(9) \quad \begin{aligned} f(x_1, x_2) = & f(x_1^0, x_2^0) + \frac{\partial f}{\partial x_1} (x_1 - x_1^0) + \frac{\partial f}{\partial x_2} (x_2 - x_2^0) \\ & + \frac{1}{2} \left\{ \frac{\partial^2 f}{\partial x_1^2} (x_1 - x_1^0)^2 + 2 \frac{\partial^2 f}{\partial x_1 \partial x_2} (x_1 - x_1^0) (x_2 - x_2^0) \right. \\ & \left. + \frac{\partial^2 f}{\partial x_2^2} (x_2 - x_2^0)^2 \right\} + \dots \end{aligned}$$

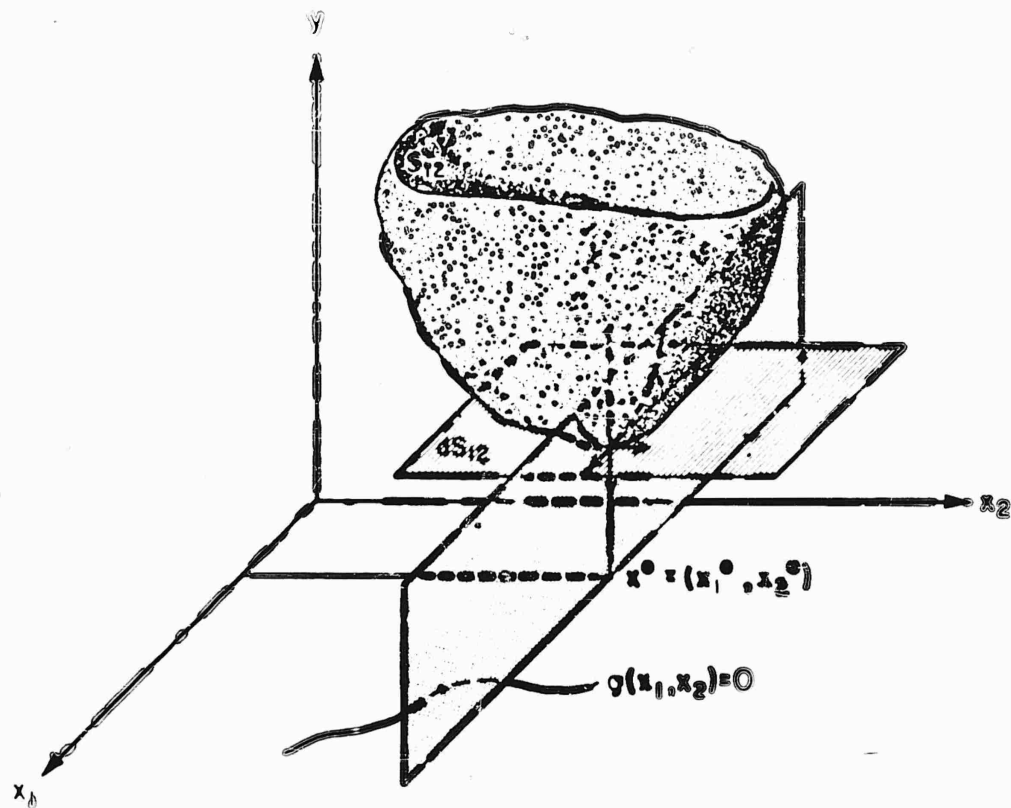


Fig. 2

By an argument similar to that used to derive (6), a sufficient condition for x^0 to be a (relative) minimum is

$$(10) \quad \frac{1}{2} \left\{ \frac{\partial^2 f}{\partial x_1^2} (x_1 - x_1^0)^2 + 2 \frac{\partial^2 f}{\partial x_1 \partial x_2} (x_1 - x_1^0) (x_2 - x_2^0) + \frac{\partial^2 f}{\partial x_2^2} (x_2 - x_2^0)^2 \right\} \geq 0$$

The tangent vectors $(dx_1, 0) = (x_1 - x_1^0, 0)$ and $(0, dx_2) = (0, x_2 - x_2^0)$ to the surface (7) at x^0 determine a 2-dimensional tangent plane dS_{12} to S_{12} . Since x^0 is to be a relative minimum, all sufficiently close neighboring points must not yield a smaller y , points on the tangent $(dx_1, 0)$, $(0, dx_2)$ being only special cases. More generally, points on any vector $d\bar{x} = (d\bar{x}_1, d\bar{x}_2)$ at x^0 which is a linear combination of $(dx_1, 0)$, $(0, dx_2)$ must also not yield a smaller y . Since $(dx_1, 0)$, $(0, dx_2)$ span or form a basis of dS_{12} , $d\bar{x}$ may be represented as

$$(11) \quad (d\bar{x}_1, d\bar{x}_2) = \cos \alpha_1 (dx_1, 0) + \cos \alpha_2 (0, dx_2)$$

where $\cos \alpha_1$, $\cos \alpha_2$ are the direction cosines of $d\bar{x}$ with respect to the local coordinate system on dS_{12} with origin at x^0 . Consequently, a strengthened necessary condition for a relative minimum at x^0 , which includes the two equations in (8) as special cases, is

$$(12) \quad \nabla_{d\bar{x}} f = \frac{\partial f}{\partial x_1} \cos \alpha_1 + \frac{\partial f}{\partial x_2} \cos \alpha_2 = 0$$

where $(\cos \alpha_1, \cos \alpha_2)$ are the direction cosines of an arbitrary tangent vector $d\bar{x}$ in dS_{12} at x^0 . $\nabla_{d\bar{x}} f$ is called the directional derivative of f in the direction $d\bar{x}$. Also, a strengthened sufficient condition for a relative minimum at x^0 is, by taking the directional derivatives of f_{x_1} , f_{x_2} again in the direction $d\bar{x}$,

$$(13) \quad \frac{\partial^2 f}{\partial x_1^2} \cos^2 \alpha_1 + 2 \frac{\partial^2 f}{\partial x_1 \partial x_2} \cos \alpha_1 \cos \alpha_2 + \frac{\partial^2 f}{\partial x_2^2} \cos^2 \alpha_2 \approx 0$$

for the chosen arbitrary tangent vector $d\hat{x}$ in dS_{12} at x^0 .

The tangent plane dS_{12} at any point \hat{x} on S_{12} is defined by the linear terms in the expansion (9), i.e.,

$$(14) \quad y - f(\hat{x}_1, \hat{x}_2) = \frac{\partial f}{\partial x_1} (x_1 - \hat{x}_1) + \frac{\partial f}{\partial x_2} (x_2 - \hat{x}_2)$$

where (x_1, x_2, y) is a point in dS_{12} , and the partial derivatives are to be evaluated at \hat{x} . To put the matter differently, if S_{12} itself is already a plane, then the expansion (9) at any point on it must be exact with only the linear terms, i.e., all higher-order terms must vanish identically. The normal to the tangent plane dS_{12} at \hat{x} , also called the gradient vector ∇f to S_{12} at \hat{x} , has components proportional to

$$(15) \quad \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, -1 \right)$$

At a relative minimum point x^0 on S_{12} , (8) holds and thus in (14) $y = f(x_1^0, x_2^0)$ identically, which is another way of saying that dS_{12} at x^0 is parallel to the $x_1 x_2$ - plane (called the base plane) and at distance $f(x_1^0, x_2^0)$ from it. At an arbitrary point \hat{x} on S_{12} , the left side of (14) need not vanish; so will the right side not also. But the right side of (14) is the same as $\nabla_{dx} f$ defined in (12) if we choose a point (x_1, x_2, y) in dS_{12} at \hat{x} such that

$$(16) \quad x_1 - \hat{x}_1 = \cos \alpha_1, \quad x_2 - \hat{x}_2 = \cos \alpha_2$$

Since $(\cos \alpha_1, \cos \alpha_2)$ represents a unit vector with respect to the local coordinate system in dS_{12} , $\nabla_{d\bar{x}} f$ is pre-

cisely the component (i.e. projection) of ∇f in the direction $d\bar{x}$.

(12) states, therefore, that at a critical point x^0 on S_{12} , the projection of the gradient vector ∇f in every direction $d\bar{x}$ in dS_{12} vanishes, and (14) shows that at a noncritical point \hat{x} of S_{12} , the projection of ∇f on dS_{12} need not vanish for all directions $d\bar{x}$ in dS_{12} . This result applies generally to spaces of dimensions greater than 2.

On the basis of the above geometric concepts, it is now possible to give an intrinsic characterization of Lagrange multipliers. Consider, for example, a constrained minimum problem of the following type: Minimize (7) subject to

$$(17) \quad g(x_1, x_2) = 0.$$

(17) defines a curve in the base plane, and minimum of f is to be sought among all points $x = (x_1, x_2)$ lying on this curve C . At any such (relative) minimum point x^0 , the directional derivative ∇f of f along the tangent to C must vanish by (12), where $\cos \alpha_1, \cos \alpha_2$ denote the components of the unit tangent dx to C at x^0 . However, (17) shows that

$$(18) \quad \nabla_{dx} g = \frac{\partial g}{\partial x_1} \cos \alpha_1 + \frac{\partial g}{\partial x_2} \cos \alpha_2 = 0$$

also at this point. Consequently, $\nabla_{dx} f$ and $\nabla_{dx} g$ must be collinear, i.e., for some scalar λ .

$$(19) \quad \nabla_{dx} f = \lambda \nabla_{dx} \epsilon$$

where dx is the tangent vector to C defined by (17). (19) is equivalent to

$$(20) \quad f_{x_1} = \lambda \epsilon_{x_1}, \quad f_{x_2} = \lambda \epsilon_{x_2}$$

which are the usual conditions derivable from differentiation of the Lagrangian function. A sufficient condition for a relative minimum at x^0 is (13) with $\cos \alpha_1, \cos \alpha_2$ being again the components of dx (the tangent to C at x^0) with respect to the local coordinate system at x^0 .

Generalization of the above geometric characterization of Lagrange multipliers to spaces of higher dimensions is immediate. Let

$$(21) \quad y = f(x_1, x_2, \dots, x_n)$$

again denote the objective function to be extremized, and

$$(22) \quad g_j(x_1, x_2, \dots, x_n) = 0 \quad (j = 1, \dots, r < n)$$

denote a set of independent side constraints. Each g_j defines a hypersurface S_j in the base plane (i.e., the (x_1, \dots, x_n) -plane in $(n+1)$ -dimensional space E^{n+1} with the last axis y). The intersection

$$(23) \quad S_{12\dots r} = \bigcap_{j=1}^r S_j$$

of these hypersurfaces in general yields an $(n-r)$ -dimensional surface in the base plane. At a critical point $x^0 = (x_1^0, \dots, x_n^0)$ on

$S_{12\dots r}$, a tangent space $dS_{12\dots r}$ generally exists with basis vectors $(dx)^1, \dots, (dx)^{n-r}$, and the directional derivative ∇f of f along each such basis vector must vanish. This says that ∇f must be orthogonal to $dS_{12\dots r}$, or ∇f lies in $dS_{12\dots r}^\perp$, the orthogonal complement of $dS_{12\dots r}$ at x^0 . But (22) shows that

$$(24) \quad \sum_{i=1}^n \frac{\partial g_j}{\partial x_i} dx_i = 0 \quad (j = 1, \dots, r)$$

also at x^0 . Hence, if $dx = (dx_1, \dots, dx_n)$ is chosen to range over the basis vectors $(dx)^1, \dots, (dx)^{n-r}$ of $dS_{12\dots r}$, (24) merely shows that each ∇g_j ($j=1, 2, \dots, r$) is also orthogonal to $dS_{12\dots r}$. But if g_j ($j=1, \dots, r$) are independent, $\nabla g_1, \dots, \nabla g_r$ would form a basis for $dS_{12\dots r}^\perp$ since

$$(25) \quad \dim dS_{12\dots r} + \dim dS_{12\dots r}^\perp = n$$

at any regular point on $S_{12\dots r}$. Therefore, for some scalars $\lambda_1, \dots, \lambda_r$ we must have

$$(26) \quad \nabla f = \sum_{j=1}^r \lambda_j \nabla g_j$$

which gives in component form,

$$(27) \quad \frac{\partial f}{\partial x_i} = \sum_{j=1}^r \lambda_j \frac{\partial g_j}{\partial x_i} \quad (i = 1, \dots, n)$$

These form a system of n equations in $n + r$ unknowns x_1, \dots, x_n ; $\lambda_1, \dots, \lambda_r$. But since (x_1, \dots, x_n) must also satisfy (22), r additional equations are to be added. Consequently, the Lagrange multipliers are merely coefficients used in expressing a certain linear dependence relation among the gradient vectors to f and g_j 's.

The sufficiency condition is also easily generalized. With respect to the basis vectors $(dx)^1, \dots, (dx)^{n-r}$ of $dS_{12\dots r}$, a typical unit tangent vector $d\bar{x}$ in $dS_{12\dots r}$ has the form

$$(28) \quad d\bar{x} = \sum_{k=1}^{n-r} (dx)^k \cos \alpha_k$$

where $\cos \alpha_1, \dots, \cos \alpha_{n-r}$ are the direction cosines

of $d\bar{x}$ with respect to

$(dx)^1, \dots, (dx)^{n-r}$. Then

$$(29) \quad \nabla_{d\bar{x}}^2 f = \nabla_{d\bar{x}} (\nabla_{d\bar{x}} f) = \sum_{i,j=1}^{n-r} \frac{\partial^2 f}{\partial x_i \partial x_j} \cos \alpha_i \cos \alpha_j \geq 0$$

together with (27) yields a relative constrained minimum at x^0 . Alternatively, if $z = (z_1, \dots, z_n)$ is any vector in the base plane, a relative constrained minimum at a point x^0 is assured by (27) and

$$(30) \quad \begin{aligned} z'Hz &= (z_1, \dots, z_n) \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \\ &= \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} z_i z_j \geq 0 \end{aligned}$$

for all z orthogonal to $\nabla g_1, \dots, \nabla g_r$. That is, for all z_1, \dots, z_n satisfying

$$(31) \quad \sum_{i=1}^n \frac{\partial g_j}{\partial x_i} z_i = 0 \quad (j = 1, \dots, r)$$

That (30) and (31) may be translated into appropriate properties of the bordered Hessian

$$(32) \quad \begin{vmatrix} 0 & \cdot & \cdot & \cdot & \cdot & 0 & \frac{\partial g_1}{\partial x_1} & \cdot & \cdot & \cdot & \cdot & \frac{\partial g_1}{\partial x_n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \frac{\partial g_r}{\partial x_1} & \cdot & \cdot & \cdot & \cdot & \frac{\partial g_r}{\partial x_n} \\ \frac{\partial g_1}{\partial x_1} & \cdot & \cdot & \cdot & \cdot & \frac{\partial g_r}{\partial x_1} & \frac{\partial^2 f}{\partial x_1^2} & \cdot & \cdot & \cdot & \cdot & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\partial g_1}{\partial x_n} & \cdot & \cdot & \cdot & \cdot & \frac{\partial g_r}{\partial x_n} & \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdot & \cdot & \cdot & \cdot & \frac{\partial^2 f}{\partial x_n^2} \end{vmatrix}$$

may also be readily established.